Multimodal fingerprint verification by score-level fusion: an experimental investigation

Gian Luca Marcialis¹, Fabio Roli, Luca Didaci*

University of Cagliari – Department of Electrical and Electronic Engineering
*University of Cagliari – Department of Pedagogical and Philosophical Sciences

Piazza d’Armi – I-09123 Cagliari (Italy)

Phone: +39 070 675 5893 – Fax: +39 070 675 5900

{marcialis, roli, luca.didaci}@diee.unica.it

Abstract. The score-level fusion approaches for fingerprint verification have been widely investigated. However, this investigation has been performed by studying each approach independently from the others, thus using different acquisition sensors, matching algorithms, fusion rules, and data sets. Due to this strong variability, the literature is lack of an experimental investigation aimed to fairly compare the various approaches. This is the scope of the present paper, from the point of view of the performance improvement especially. In our opinion, this investigation can allow to confirm state-of-the-art results by further experimental evidences.

1 Introduction

In recent years, automatic fingerprint verification systems have been proposed to certify the identity of a person [1-11]. Advantages of fingerprints are well-known in terms of uniqueness and acceptability. Moreover, they are difficult to steal and reproduce. Accordingly, various matching algorithms based on fingerprints have been proposed so far [1]. However, a system based on single sensor and/or matcher did not allow a large coverage of the user population. In addition, no single matcher has been able to meet the stringent

¹ Corresponding author
accuracy requirements of real applications. Therefore, recent works argued the use of information fusion in terms of score combination to improve the current performance of the fingerprint verification systems.

According to the recent paper by Prabhakar and Jain [3] and Ross et al. [12], the score-level fusion techniques in fingerprint verification can be classified into four scenarios:

1. Fusion of multiple matching algorithms (multi-algorithmic systems): two or more matching algorithms are applied to a unique fingerprint impression.

2. Fusion of multiple impressions: the same matching algorithm is applied to two or more impressions of the same finger (or different fingers).

3. Fusion of multiple sensors: the same matching algorithms is applied to two or more images provided by two or more fingerprint sensors.

4. Fusion of fingerprints and other biometrics: the score provided by the fingerprint matcher is combined with that provided by other biometric matchers (e.g. a face matcher).

The most widely investigated scenario is that of the fusion of multiple matching algorithms (e.g. [3, 4, 9, 14]). The rationale behind this approach is that scores obtained by different matching algorithms provide different information which can be exploited by their combination. In [3] the authors investigated four different matching algorithms and presented a matcher selection scheme to provide the best matchers for combination. Fusion of the matching scores is performed by sum, product rules and the estimate of the joint probability of impostor and genuine users classes for performing the Neyman-Pearson test (null hypothesis: the pattern is impostor). The fusion of a minutiae-based matcher and a texture-based matcher, namely, String and Filter, gave the best results. In [4] the authors improved the texture-based algorithm presented in [3] and combined it with the String matcher by Linear Discriminant Analysis (they treated the couple of matching scores as a novel pattern to be submitted to LDA). In [9] we reported further experimental evidences about the
performance improvement achievable by score-level fusion of minutiae-based and texture-based algorithms by performing a weighted averaging of the scores. Further experimental confirmations are reported in [14] when more than two matchers, among the ones submitted to the Fingerprint Verification Competition on 2004, are fused with several approaches.

To the best of our knowledge, the second scenario has been investigated in [3] and [13] only. It is reasonable to argue that adding the information related to each impression can contribute to reduce the intra-class variation, thus giving a more reliable classification of the subject. In [3] the authors showed that by using the Neyman-Pearson approach multiple impressions-based systems can improve the performance with respect to that of the best individual matcher. Similar results have been obtained in [13] using score-level averaging.

Finally, the third scenario has been proposed by the authors in [5]. It is well-known that the information obtained by different acquisition sources can be complementary. Therefore, combining this information can provide a more reliable matching score and thus a better classification of the subject. In [5], the matching scores were provided by an optical and a capacitive sensor with the String matching algorithm. The investigated combination rules were the sum, product, maximum, minimum of the scores, and a perceptron-based fusion rule trained to maximise the separation of genuine user and impostor classes [9]. Even in this case, fusion definitely improved performances, and also exhibited the property of “recovering” patterns correctly classified by only one of the individual matchers. This result suggested that multi-sensor systems can be more robust than single sensor ones in classifying fingerprint images affected by dryness or moisture due to the different physical principle of the adopted acquisition sources.

The last scenario is not related to fingerprint matchers only. Since the focus of this paper is the combination of fingerprint matchers, namely, the investigation of the first three scenarios,
we disregard it in the following. In the following we refer to the three investigated scenarios with the term “multimodal fingerprint verification”.

It is worth noting that, although all scenarios were investigated, generally this has been done by using “home-made” data sets. This problem has been recently raised by Norman and Poh in [11]. Therefore, the actual performance improvement achievable by these approaches is difficult to appreciate, because no common protocol was applied to each experiment. With regard to this issue, Table 1 summarises the main achievements of the state-of-the-art. The first column specifies the application scenario and the related reference, the columns from the second one to the fourth one indicates the acquisition sensor, the individual algorithms, the proposed fusion rule, the fifth column gives some information about the used data set, and the sixth and seventh columns point out the best individual result and the best fusion result. We used the 1%FAR value since it represents a very stringent performance requirement.

It is difficult to compare Table 1 results due to the strong variability of the used data sets, fusion rules, matching algorithms. As an example, in [3] results using the Neyman-Pearson approach are reported but no comparison is made using standard fusion rules. In principle, Table 1 points out the need of a comparison performed on benchmark data sets. On the other hand, Table 1 points out that there is need to have further evidence on the above scenarios: as an example, results reported in [3] show that multiple matchers fusion and multiple impressions fusion exhibit similar performance values, whilst it could expected that using multiple impressions should add discriminant information, as pointed out in [13].

In this work we tried to fill such gaps by comparing the three scenarios in terms of the verification performance on widely adopted and publicly available benchmark data sets.

The rest of the paper is organised as follows. In Section 2 we describe the algorithms and the sensors used for fingerprint matching. Experimental results comparing the performance of the
various information fusion approaches are reported in Section 3, and conclusions are drawn in Section 4.

Table 1. Performances of mono-sensor systems (optical and capacitive) and various score-level fusion techniques referred to the three investigated Scenarios (multi-algorithmic fusion, multi-impression fusion and multi-sensor fusion).

<table>
<thead>
<tr>
<th>Scenario [Ref.]</th>
<th>Sensor</th>
<th>Combined algorithms</th>
<th>Fusion rule</th>
<th>Data set</th>
<th>Best individual (1% FAR)</th>
<th>Fusion Result (1% FAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-algorithmic [3]</td>
<td>Optical</td>
<td>Minutiae-based (String, Dynamic, Hough) and Filter</td>
<td>Neyman-Pearson (NP), Mean, Product</td>
<td>Home-made: 668 classes, 4 impressions per class</td>
<td>10.0% (String)</td>
<td>1.9% (NP), 2.3% (mean), 3.3% (product)</td>
</tr>
<tr>
<td>Multi-algorithmic [4]</td>
<td>Capacitive</td>
<td>String and Filter</td>
<td>LDA</td>
<td>Home-made: 640 classes, 4 impressions per class</td>
<td>23.0% (Filter)</td>
<td>10.0%</td>
</tr>
<tr>
<td>Multi-algorithmic [9]</td>
<td>Optical</td>
<td>String and Filter</td>
<td>Perceptron using Fisher distance, Mean rule</td>
<td>FVC2000-DB1: 100 classes, 8 impressions per class</td>
<td>2.9% (String)</td>
<td>1.8%, 1.7% (mean)</td>
</tr>
<tr>
<td>Multi-algorithmic [14]</td>
<td>Optical, Thermal</td>
<td>Algorithms from competitors to FVC 2004</td>
<td>Sum rule, SVM, Dempster-Shafer</td>
<td>FVC 2004: 4 data sets, 100 classes, 8 impressions per class</td>
<td>Around 1.0%</td>
<td>Around 0.5% (sum)</td>
</tr>
<tr>
<td>Multi-impressions [3]</td>
<td>Optical</td>
<td>String and Filter</td>
<td>Neyman-Pearson (NP)</td>
<td>Home-made: 668 classes, 4 impressions per class</td>
<td>10.0% (String)</td>
<td>2.0%</td>
</tr>
<tr>
<td>Multi-impressions [13]</td>
<td>Optical</td>
<td>String</td>
<td>Neyman-Pearson (NP), Mean rule, majority voting</td>
<td>Home-made: 827 classes, 8 impressions per class</td>
<td>7.5% (home-made)</td>
<td>2.5% (mean)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FVC 2002-Db1/2: 110 classes, 8 impressions per class</td>
<td>15.0% (FVC-Db1), 8.0% (FVC-Db2)</td>
<td>7.5% (mean), 6.0% (mean)</td>
</tr>
<tr>
<td>Multi-sensor [5]</td>
<td>Optical and Capacitive</td>
<td>String</td>
<td>Perceptron using Fisher distance, mean rule</td>
<td>Home made: 120 classes, 10 impressions per class</td>
<td>5.3% (String – Optical)</td>
<td>3.0% (Perceptron-FD), 5.0% (mean)</td>
</tr>
</tbody>
</table>
2 Score-level fusion in fingerprint verification: the investigated algorithms

We used fixed and trained rules as fusion: the first ones are based on simple average and product of matching scores, the second ones are represented by the well-known Linear Discriminant Analysis (LDA) based on the Fisher distance maximization criterion [4], and standard perceptron-based fusion maximizing the likelihood [15]. Whilst average and product are quite simple, LDA and perceptron-based fusion rules are called “trainable”, such that the final score is obtained as follows:

\[ s = w_0 + w_1 s_1 + w_2 s_2 \]  

(1)

In Eq. (1), \(s_1\) and \(s_2\) are matching scores provided by the individual algorithms. The set of real values \(\{w_0, w_1, w_2\}\) is computed by training the related algorithm on a set of samples (named “training set”) in order to optimize the objective function, namely, Fisher distance or likelihood [4, 15]. An additional step of score normalization is given by:

\[ s_{\text{norm}} = \frac{1}{1 + \exp(-s)} \]  

(2)

This is aimed to normalize the score values into [0, 1] interval, according to the definition of matching score. Eq. (2) is called sigmoid or logistic function.

Scenarios 1, 2, 3 have been investigated using such fusion rules, which fit quite well the hypothesis of sources “independence”. Although previous works showed that average and product of scores can be sufficient to definitely improve performances without requiring parameters setting by additional training sets [3, 5, 9, 10, 12, 13], we added LDA and perceptron-based fusion rules which showed good results in several of the investigated scenarios [4, 9, 16].

With regard to the acquisition sources, we selected an optical sensor and a capacitive sensor, because they are used extensively to test fingerprint matching algorithms [6]. Moreover, their
physical principles of acquisition are so different that it is reasonable to expect that “fusion” of the information provided by each of them can be effective [5, 6].

After the image acquisition, feature extraction and matching algorithms are applied separately to the each fingerprint image. Then, matching scores are fused with the selected rules. Finally, a threshold-based decision is used for the genuine/impostor classification.

We used two kind of fingerprint matchers: the first one is based on minutiae points, the second one is based on the FingerCode.

The so-called “minutiae” points correspond to the bifurcations and the terminations of the ridge lines [1]. A review about minutiae-based approaches to fingerprint verification recently appeared in [8]. In this work, minutiae have been extracted from the skeletonised fingerprint images obtained by the commonly used enhancement and binarization steps. A further post-processing phase has been necessary in order to delete some spurious minutiae [1, 5, 8]. Two fingerprints are compared by the respective set of minutiae points, so generating a matching score. This score is proportional to the number of minutiae couples of the two fingerprints which can be considered as “aligned”. The most used algorithm based on this approach is called “String” [2-5]. Briefly:

1. Each minutia is characterized by its coordinates in the x-y axis and its orientation with respect to x (but it can be considered another reference axis as well).

2. A “minutiae distance” is defined as weighted sum of the Euclidean distance and the orientation distance among two minutiae.

3. Let \( A \) be the template minutiae set.

4. Let \( B \) be the input minutiae set.

5. For each minutia \( a \in A \),
   a. For each \( b \in B \), \( a \) is aligned to \( b \). Alignment is performed by performing a roto-translation of \( b \) on \( a \) according to the related coordinates and orientations.
b. After this alignment, $a$ and $b$ match perfectly, that is, the “distance”. Let $AL(a, b) = \{ (a_i, b_i), a_i \in A, b_i \in B : \text{aligned}(a_i, b_i) = true \}$ be the set of other couples of aligned minutiae. $a_i$ and $b_i$ are considered as aligned on the basis of the above “minutiae distance” not exceeding a certain fixed threshold.

6. The value $\max_{a,b} \{ |AL(a, b)| \}$ is converted to the matching score by the formula:

$$\text{score} = \frac{(\max_{a,b} \{ |AL(a, b)| \})^2}{|A| \cdot |B|}$$

(3)

The FingerCode-based systems use a feature vector which describes the “shape” of the ridge lines along the fingertips [3-4]. The FingerCode is obtained as follows:

1. Tessellate the fingerprint image by several blocks of a fixed size (11x11 pels).

2. For each block, compute the response of a set of Gabor filters, which have orientation-selective characteristics, for each part of the tessellation. Gabor filters expression is given in [4]. Parameters of Gabor filters are the ridge frequency, which we set as 4 pixels according to the values given in [4] for 500 dpi images, and the orientation. We use eight possible orientations for each block.

3. For each filtered block, compute the variance among pixels gray-levels. Accordingly, a feature vector is obtained, made up of $eight \times (#blocksPerImage)$ values.

4. The comparison among fingerprints is performed by evaluating the Euclidean distance among the respective FingerCodes and converting it in a matching score. We added to the algorithm described in [4] the following modification.

   a. We computed two matching scores: the first one is due to the input FingerCode obtained by the best alignment with respect to the template fingerprint computed by String (as described in [4]);

   b. the second one is due to the FingerCode obtained with no alignment. Then, we selected the maximum score among these. The scheme is shown in Figure 1.
Our choice has been due to the fact that the “best alignment” given by String was actually a “bad” alignment due to the failure of minutiae matching, in certain cases. Therefore, the matching score produced by the comparison of two FingerCodes was too low. Our choice allowed to avoid some of these unfavourable cases. Experiments showed that our modified algorithm works better than the original one. We refer to this algorithm as “Filter” in the following.

Adopted “String” and “Filter” algorithms version are publicly available, and can be obtained by contacting the authors.

3 Experimental results

3.1 The data sets

For our experiments we used the FVC-2000 DB1 and DB2 data sets, proposed for the international fingerprint verification competition [7]. These data sets are made up of 800 fingerprints images belonging to 100 subjects, acquired with an optical sensor (DB1) and a capacitive sensor (DB2). They can be used also as “multi-sensor” data sets because fingerprint images have been taken from the same subjects [7]. In general, none of the FVC2000 data sets were collected with the aim of using them in a multi-sensor fusion context. However, FVC2000-Db1 and Db2 can be considered as an exception to that rule, for the following reasons: firstly, they have been collected from the same user population, which means that the users of Db1 and Db2 are the same, as written in [7], official report of FVC2000; secondly, before performing our experiments, we manually and carefully verified that each image related to the same finger index for DB1 and DB2 correspond. We did this for all users in the data sets. In other words, we verified that the i-th user of FVC2000 DB1 is the same i-th user of FVC2000 DB2. Therefore, we can confirm without doubt that the above
data sets can be used as multi-sensorial data sets. Figure 2 shows DB1 and DB2 images for twenty-five randomly chosen subjects with same user identifier. It is easy to see, even by visual inspection, that optical and capacitive images refer to the same fingerprint, thus, to the same user. To the best of our knowledge, it is the first time that the all three scenarios described in Section 1 are experimentally investigated on such a publicly available data set.

Images have been taken at 500 dpi. Further details on FVC2000 data sets can be found in [7]

### 3.2 Experiments plan

In our experiments, the FVC protocol was adopted [7]:

- For each mono or multi-sensor verification algorithm, two sets of scores were computed from the so-called “set b” of FVC data sets. The first one is the so called genuine scores set $G$, created by matching all possible couples of equal fingerprints. A fingerprint with score from the set $G$ obviously belongs to the “genuine user” class. The second one is the impostor scores set $I$, created by matching fingerprints of different persons. A fingerprint with score from the set $I$ belongs to the “impostor” class. The numbers of genuine and impostor comparisons, that is, the sizes of sets $G$ and $I$, were 2,800 and 314,000, respectively.

- Performances were assessed and compared in terms of:
  - the so called Equal Error Rate (EER) measure, that is, the verification error rate evaluated at the threshold value for which the False Acceptance Rate (percentage of accepted impostors) is equal to the False Rejection Rate (percentage of rejected genuine users).
  - the so called 1%FAR and 1%FRR measures. 1%FAR is the FRR when the FAR is fixed to 1%. On the contrary, 1%FRR measure assesses the FAR when FRR=1%. These performance measures are used for assessing verification systems under the stringent requirements of many real applications. For example, 1%FRR
performance measure allows assessing the security degree of the system (i.e., the FAR of the system) when only 1% of genuine users can be rejected.

With regard to LDA and perceptron-based fusion rules, their parameters have been estimated on the “set a” of FVC data sets.

For sake of clarity, due to the large amount of experiments performed, we only reported the best and worst results obtained by fusion in each of investigated Scenarios.

3.3 Results

Table 2 reports best and worst values of EER, 1%FAR and 1%FRR performance measures for the mono-sensors systems, and the various score-level fusion techniques by using String and Filter as matching algorithms and optical and capacitive sensor for image capturing. The various scenarios are presented by increasing their intrinsic complexity according to Section 1.

On the FVC data sets, performances of optical sensor are substantially lower than the ones of the capacitive sensor, although usually an optical sensor is claimed better than capacitive ones when using a minutiae-based matching algorithm as also showed in Table 1 [4, 5]. This further points out that no sensor can be claimed better than the others for all kinds of input fingerprints and noise in sensed data.

With regard to matchers adopted, they strictly implement state-of-the-art algorithms as String and Filter. According to [7], official report of FVC2000, the performance of adopted String algorithm falls in the middle of the rank which can be found in Table III (p 16) for Db1 and in the fifth position (on eleven participants) of Table IV (p. 17). With regard to Filter algorithm, its performance falls in the middle of those ranks (pp. 16-17). We are allowed to compare the above results because we adopted the same FVC protocol in our paper. Therefore, adopted matchers can represent on average the performance achievable on FVC2000 DB1 and DB2 data sets.
The fusion of multiple matchers is able to improve the performance with respect to that achieved by using a single matching algorithm, and confirm results reported by other authors even in terms of the amount of the average improvement which can be observed by comparing Table 1 (six-seventh columns) and Table 2 (third column). On the other hand, its performance is considerably lower than that of the multi-sensor fusion. The performance difference is very high in worst fusion results. The amount of the improvement is strongly superior than that reported in [5] (Table 1, last row), so increasing the relevance of our results since they are reported on benchmarking data sets. Multi-sensor fusion exhibits a performance globally comparable with that of the multi-impressions fusion, even if it allows good 1%FAR values even in worst case (Table 2, third column). As pointed out in Section 1, this is very likely to happen, due to the use of multiple information sources coming from multiple images. Therefore, both strategies can be used for improving performances as well, and their contribute is better than that coming from fusion of multiple matchers. This result is in agreement with the hypothesis that multiple impressions add discriminant information [13], and differs from that presented in [3] and reported in Table 1 (second and fifth row). Improvements pointed out by trained rules are in the case of multi-algorithmic and multi-sensors scenarios (Scenarios 1 and 3, respectively). This could be motivated by the fact that these rules are less conditioned by score dynamic provided by matchers, which lead to different genuine and impostors distributions, since they “tune” parameters over a training set of scores (obviously representative of expected unknown samples). Thus they are able to manage those dynamics better than fixed rules. On the other hand, in the case of multiple-impressions scenario, trained rules appear to be superfluous, because scores derive from the same genuine users and impostors distributions. It is also worth noting that, if matcher is too weak (Filter), there is not strong difference among fixed and trained rules for Scenario 2.
Worst results are quite similar independently on the fusion rule adopted. With regard to Scenarios as 1 and 3, score normalization steps are required [18] in order to allow a fixed rule to perform at best, but, at the state-of-the-art, they must be designed appropriately for adopted matchers and fusion rules.

On the basis of the above results, it is possible to conceive an automatic verification system which combines the advantages of multi-algorithmic and multi-sensor systems. It can be hypothesised that the first ones are able to improve performances by extracting complementary feature from the same fingerprint image, the second ones are able to improve performances by exploiting the complementarity of different acquisition sources, as the optical and the capacitive ones.

Accordingly, fingerprints could be acquired by the optical and capacitive sensors, as standard multi-sensor systems. For each sensor, fingerprint images could be processed by a multi-algorithmic system which performs matching. Each couple of matching scores \( \{s_{o1}, s_{o2}\} \) and \( \{s_{c1}, s_{c2}\} \) is fused. The resulting couple of matching scores \( \{s_o, s_c\} \) is combined with the same rules. The best performance of such a “multi-level” system is showed in Table 3. This performance has been obtained by combining the best multiple matchers systems of Table 2, that is, the systems using the product rule.

Reported results confirmed our hypothesis about the complementarity of multi-algorithmic and multi-sensor systems. It is particularly relevant that high security parameters 1%FAR and 1%FRR approach to the same value. This means that no significant variation is exhibited by changing the operational point conditions (i.e. requiring a more stringent constraint in accepting a client or rejecting an impostor). Therefore, the impact of estimation errors of the acceptance threshold is strongly reduced.

Results provided in Tables 2 confirm the hypothesis of the complementarity of information coming from multiple impressions and sensors, by better dealing with intra-class variation
due to distortion, pressure and, in general, the subject cooperation. But the further improvement of Table 3 allows us to add that the contribute of multi-sensor and multi-impressions are not the same. In order to explain this point of view, we reported in Figure 3 some images from the same subject acquired under different moisture state. In these images, we artificially increase the moisture (or dryness) of the finger skin. Fingerprint sensors adopted are the same we already used in [5]. It can be noticed that images provided by these sensors are strongly different. This supports the hypothesis that multiple sensors can help in the case that subject co-operates but, for several reasons, his emotional or health state is altered. It is worth considering that if the finger is too dry, it is not useful to acquire multiple impressions using only one sensor, e.g. an optical one, because the problem is intrinsic in the physiological state of the finger skin.

Table 2. Performances of mono-sensor systems (optical and capacitive) and various score-level fusion techniques on FVC2000 DB1 and DB2 data sets by using the String and Filter matching algorithms from optical and capacitive sensor images in terms of EER, 1%FAR and 1%FRR. Percentage values are given.

<table>
<thead>
<tr>
<th>Individual systems</th>
<th>EER</th>
<th>1% FAR</th>
<th>1% FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical – String</td>
<td>10,5</td>
<td>22,4</td>
<td>83,5</td>
</tr>
<tr>
<td>Capacitive – String</td>
<td>5,4</td>
<td>9,6</td>
<td>49,3</td>
</tr>
<tr>
<td>Optical – Filter</td>
<td>13,7</td>
<td>46,9</td>
<td>82,0</td>
</tr>
<tr>
<td>Capacitive – Filter</td>
<td>12,6</td>
<td>57,9</td>
<td>65,3</td>
</tr>
</tbody>
</table>

Best Fusion Results

| Scenario 1: Multi-algorithmic (Optical – Product) | 8,5  | 17,2 | 74,3 |
| Scenario 1: Multi-algorithmic (Capacitive – Perceptron) | 4,0  | 9,8  | 11,4 |
| Scenario 2: Multi-impressions (Optical – String – Mean) | 5,3  | 10,4 | 41,6 |
| Scenario 2: Multi-impressions (Capacitive – String – Mean) | 2,2  | 3,2  | 8,5  |
| Scenario 3: Multi-sensors (String – Mean) | 2,5  | 3,9  | 13,3 |
| Scenario 3: Multi-sensors (Filter – LDA) | 4,7  | 15,3 | 10,3 |

Worst Fusion Results

| Scenario 1: Multi-algorithmic (Optical – LDA) | 12,5 | 34,5 | 84,9 |
| Scenario 1: Multi-algorithmic (Capacitive – Mean) | 7,3  | 20,7 | 54,9 |
| Scenario 2: Multi-impressions (Optical - Filter – Product) | 12,2 | 37,1 | 65,5 |
| Scenario 2: Multi-impressions (Capacitive - Filter – Mean) | 4,5  | 16,3 | 11,3 |
| Scenario 3: Multi-sensors (String – Product) | 2,7  | 4,5  | 21,0 |
| Scenario 3: Multi-sensors (Filter – Product) | 9,5  | 28,9 | 57,2 |
Table 3. Performances of the multi-level system based on multi-sensor fusion of multi-algorithmic scores provided by the optical and capacitive systems in terms of EER, 1% FAR, and 1% FRR. Percentage values are given.

<table>
<thead>
<tr>
<th>Multi-level fusion</th>
<th>EER</th>
<th>1% FAR</th>
<th>1% FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>by Product</td>
<td>1.4</td>
<td>2.0</td>
<td>1.8</td>
</tr>
<tr>
<td>by Mean</td>
<td>1.5</td>
<td>2.4</td>
<td>1.7</td>
</tr>
</tbody>
</table>

In order to investigate the behaviour of multi-sensor fusion in cases like those of Figure 3, we collected a data set, called “DIEE-Extreme”, made up of couples of optical and capacitive images by varying the degree of pressure and dryness of fingerprints. Images were acquired by the optical sensor Biometrika FX2000 (312x372 pixels, 569 dpi) and Precise Biometrics MC100 (250x250 pixels, 500 dpi). Images were collected by altering the physiological state of the finger skin in terms of dryness, so generating generated twenty couples of fingerprint images per identity. The number of considered identities, that is, different fingerprints, is 42. This data set is considerably more difficult due to the images reproducing strongly dry and moist fingers. The joint use of FVC and DIEE-Extreme allows us to draw some reliable conclusions about score-level fusion of multiple matchers. Images of Figure 3 are examples of fingerprint couples from this data set. A preliminary and easy version of this data set has been used in [5]. DIEE-Extreme is publicly available by contacting the authors. This data set has been recently adopted for studying quality measures effectiveness in optical and capacitive fingerprint images [17].

Table 4 shows the “recovery rate” of multi-sensor fusion, that is, the percentages of fingerprints correctly recognized by only one of the sensors and recovered by multi-sensor fusion. In other words, for each sensor and for the genuine and impostor classes, the values in Table 4 indicate the percentages of fingerprints in the test set that were correctly recognized thanks to multi sensor fusion. The recovery rate has been computed at the EER point estimated on the DIEE-Extreme data set when using String and Filter algorithms. Even in this case, we performed experiments with fixed and trained rules but reported only best results.
given using fixed rules. This result can be easily explained by the fact that training conditions were strongly different from test conditions, thus parameters computed by perceptron and LDA matchers were not reliable.

Table 4. Percentages of genuine and impostor fingerprints correctly recognized by only one of the considered sensors and recovered, i.e., correctly recognized, by the different fusion rules (the term “recovery” rate is used to indicate these percentages). For each sensor, recovery rates are computed with respect to the maximum number of wrongly recognized genuine patterns (second and fourth columns) and impostor patterns (third and fifth columns) in the test set.

<table>
<thead>
<tr>
<th></th>
<th>Recovery Rate Optical Sensor</th>
<th>Recovery Rate Capacitive Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genuine</td>
<td>Impostor</td>
</tr>
<tr>
<td>String</td>
<td>Product</td>
<td>74.50</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>77.56</td>
</tr>
<tr>
<td>Filter</td>
<td>Product</td>
<td>26.60</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>68.39</td>
</tr>
</tbody>
</table>

Results become particularly significant when we observe that, using e.g. the String algorithm, the worst performance of single sensor systems on the Extreme data set is that of the capacitive sensor which exhibits a FAR equal to 20.38% and a FRR equal to 33.19%, versus an expected EER of 10.31%. Even the best system, the optical-based one, exhibits a very high FRR (12.78%) versus an expected EER very low (around 1%). This confirms our claim of the introduction: it is very difficult to correctly classify fingerprints by a mono-sensor systems in case of low degree of cooperation or altered physiological characteristics of the fingers, which can arise in function of the emotional state of the subject and the environmental conditions. On the contrary, Table 4 shows that multi-sensor fusion is able to recover many
wrong classifications coming from the optical or the capacitive sensors when finger skin dryness is altered. Results are relevant for the capacitive sensor especially, because, in this case, it exhibit a small acquisition surface. This is commonly considered a limitation for capacitive sensors in the literature [1] (worth noting, these improvements have been obtained by using very simple rules as product and mean).

5 Conclusions

In this paper, we reported an experimental investigation on various approaches for fusing multiple matchers based on the fingerprint verification. Experiments had been performed with FVC2000 and DIEE-Extreme benchmark data sets, with fixed and trained fusion rules. Reported results pointed out that multi-sensor fusion provided a verification accuracy notably higher than the one of the best single-sensor fingerprint matcher and also outperformed the fusion of multiple single-sensor matchers. Moreover, it exhibited a performance comparable with that of mono-sensor systems based on fusion of multiple impressions of the same finger. We also investigated the multi-level fusion of verification systems. It is worth noting that the most relevant improvements have been obtained using simple and fixed rules, as product or mean of scores, thus confirming that independence assumption of matching score for the investigated Scenarios is a good choice.

Due to the additional deterrence property of multi-sensor systems (deceiving two different sensors instead of one), the decreasing cost of optical and capacitive sensors, the simplicity with which it is possible to integrate multi-algorithmic and multi-sensor fusion, and the strong performance improvement achievable using both approaches, designing fingerprint verification systems using different sensors resulted as the best approach in multi-modal fusion of fingerprint matchers.
References


Fig. 1. The algorithm for computing the matching score by FingerCode extraction.
Fig. 2. Twenty-five randomly chosen users from FVC2000 DB1 and DB2 data sets with the same user identifier (selected raster ids are 2, 8, 15, 21, 24, 25, 26, 27, 29, 30, 31, 39, 42, 50, 51, 55, 68, 85, 87, 88, 91, 92, 94, 97, 99). They represent couples of images captured by using the optical and capacitive sensors. It is worth noting that i-th user in the FVC2000 DB1 corresponds to the i-th user in the FVC2000 DB2 data set.
Fig. 3. Example of fingerprint images acquired by optical (left) and capacitive sensor (right) used in [5]. (a) The same fingerprint characterised by high dryness. (b) The same fingerprint characterised by high moisture.